



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE CIENCIAS**

**Proceso de identificación de cúmulos de riesgo  
en pólizas de vida individual que no pudieron  
detectarse de manera automática.**

**REPORTE DE TRABAJO  
PROFESIONAL**

**QUE PARA OBTENER EL TÍTULO DE:**

**A C T U A R I O  
P R E S E N T A:**

JUAN GUILLERMO GARZA GALINDO FLORES



**T U T O R:**  
DR. JOSÉ DE JESÚS GALAVIZ CASAS  
2011



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## Hoja de datos del jurado

### 1. Datos del alumno

Garza Galindo  
Flores  
Juan Guillermo  
54 46 21 55  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Actuaría  
098516832

### 2. Datos del tutor

Dr.  
José de Jesús  
Galaviz  
Casas

### 3.- Datos del sinodal 1

Dra  
Elisa  
Viso  
Gurovich

### 4.- Datos del sinodal 2

Act  
Jaime  
Vázquez  
Alamilla

### 5.- Datos del sinodal 3

Dra  
Ana  
Meda  
Guardiola

### 6.- Datos del sinodal 4

LCC  
Francisco Lorenzo  
Solsona  
Cuz

### 7. Datos del trabajo escrito

Proceso de identificación de cúmulos de riesgo en pólizas de vida individual que no pudieron detectarse de manera automática

27 p  
2011

## Agradecimientos

La dedicación y el esfuerzo que conlleva la realización de este reporte de trabajo esta llena de circunstancias e involucra a muchas personas. Desde la fuente de vida que generó mi existir hasta la fuente de vida que me alimenta día a día y maximiza con los más sublimes deleites personales y espirituales mi vida actualmente. Esta fuente de vida es el amor.

Dedico el presente trabajo con mucho cariño y amor a mi hermosa y amada esposa Michel Huesca Lobato, mi dulce y amado pez, por quien he tenido el placer de conocer, contemplar, disfrutar y deleitarme con las dulzuras del amor y a quién agradezco su insistencia para culminar la realización de este reporte de trabajo. ¡TE AMO PEZ!

A mi querida y dedicada madre, Alicia Laura Flores Enríquez, y a mi paciente y bondadoso padre Guillermo Bosco Garza Galindo; mis adorados padres, quienes por su amor llegué a este bello mundo y por el amor que me ofrecieron como hijo crecí, recibí educación y formé mi carácter y forma de ser. Me he sentido amado incondicionalmente, muchas gracias por el apoyo que me dieron en el esfuerzo de hacer y terminar una carrera profesional, al cabo de muchos y enormes sacrificios para ustedes.

A mis queridos hermanos, Luis Manuel y Rodrigo con quienes compartí los mayores y mejores momentos de mi niñez y adolescencia y quienes complementaron mi crecimiento y madurez con su respeto y hermandad ante todas las dificultades.

A todos mis amigos y familiares que de diversas maneras han cooperado para mi formación, con diversión y ayuda. Por su amistad y afecto, el transcurso de los días se facilita y se goza.

A todos los maestros que desde pequeño me iniciaron en el aprendizaje académico, social e intelectual y que posteriormente en su enseñanza y disciplina encontré los conocimientos necesarios para elegir mi carrera, enamorarme de ella y desarrollarla a nivel profesional.

Muchas gracias a todos por haber contribuido a mi desarrollo, pasando por las experiencias y educación que me han desenvuelto como una persona ética, de principios y valores humanos.

**Abril 2011**

# ÍNDICE

<b>1. Introducción</b> .....	<b>3</b>
1.1 El reaseguro .....	3
1.2 Marco legal .....	3
1.2.1 Entidades regulatorias .....	3
1.2.2 Contratos de reaseguro .....	3
1.3 Objetivos y limitaciones del reaseguro .....	4
1.4 Cúmulos de riesgo .....	5
1.5 Problemas en la identificación de cúmulos .....	5
Reseña .....	6
<b>2. Descubrimiento de los errores en los cúmulos de vida individual</b> .....	<b>7</b>
2.1 El origen de la controversia .....	7
2.2 Extensión del tema a toda la cartera .....	7
2.3 Causales de error en nombre .....	9
<b>3. Proceso de identificación de cúmulos no detectados</b> .....	<b>9</b>
3.1 Objetivos y alcances .....	9
3.2 Diseño general .....	9
3.3 El problema con las fuentes de información .....	10
3.4 Fase 1. Base de datos de entrada .....	11
3.5 Fase 2. Depuración de la información .....	11
3.5.1 Agrupación: .....	11
3.5.2 Separación de cartera: .....	12
3.5.3 Flujo de información: .....	12
3.5.4 Proceso de estandarización de nombres: .....	13
3.6 Fase 3. Determinación del mismo asegurado .....	16
3.6.1 Algoritmo de manera exacta: .....	16
3.6.2 Algoritmo probabilístico: .....	17
3.6.3 Matriz de coincidencias .....	18
<b>4. Resultados del proceso y conclusiones</b> .....	<b>20</b>
4.1 Revisión de la base de coincidencias .....	20
4.2 Errores detectados en la fecha de nacimiento .....	21
4.3 Efectos de la corrección en cúmulos .....	22
4.4 Estrategia de reaseguro resultante .....	24
4.5 Comentarios finales .....	24
4.6 Experiencia personal .....	25
<b>Anexo único - Estadísticas de fechas de nacimiento</b> .....	<b>27</b>

## 1. Introducción

### 1.1 El reaseguro

Las teorías de la administración de riesgos coinciden en que existen 4 actividades que el ser humano ha adoptado a lo largo de su historia para hacer frente a los riesgos que se le presentan, a saber: la aceptación, previsión, confrontación y transferencia del riesgo.

Las compañías aseguradoras capitalizan y hacen posible de manera masiva el punto de mayor trascendencia dentro de esta taxonomía: la transferencia del riesgo.

El que sea factible para cualquier persona o empresa trasladar sus riesgos a entidades de mayor alcance a juicio del interesado, es lo que fundamenta el surgimiento y existencia de la industria de los seguros.

Así como las personas, empresas o gobiernos transfieren riesgos a las compañías de seguros, éstas a su vez pueden (y en ciertas ocasiones, deben) transferir parte de los riesgos asumidos a las compañías reaseguradoras. Existe un tercer nivel de transferencia de riesgos el cual es conocido en el gremio como la retrocesión. De esta manera, un riesgo muy grande puede ser transferido de un cliente inicial a una aseguradora, ésta lo cede en una parte a una reaseguradora y finalmente ésta transfiere otra porción a una retrocesionaria.

Cabe mencionar que en una relación de reaseguro no existe vínculo alguno entre el asegurado y el reasegurador ni entre las aseguradoras para con las retrocesionarias. Además, la relación se basa y hace referencia en todo momento al precepto de la *absoluta buena fe*, difundido en el gremio y donde ambas partes básicamente extienden su confianza entre sí.

### 1.2 Marco legal

#### 1.2.1 Entidades regulatorias

Dado que los seguros tienen que responder ante eventos fortuitos, cuya volatilidad puede llegar a ser muy alta, en todos los países se han creado instituciones (generalmente dependencias gubernamentales) que regulan, vigilan y autorizan las operaciones, resultados y solvencia financiera de las compañías que forman la industria de seguros, en aras de ver por el bienestar de los asegurados.

Por tal motivo, las empresas de la industria de los seguros deben acatar las normas que dicten dichas entidades regulatorias.

En México, la CNSF (Comisión Nacional de Seguros y Fianzas) es la autoridad encargada de estas actividades y depende de la SHCP (Secretaría de Hacienda y Crédito Público).

#### 1.2.2 Contratos de reaseguro

Para que una aseguradora pueda transferir los riesgos de sus pólizas a los reaseguradores, debe existir formalmente un contrato de reaseguro entre ambas compañías. La aseguradora en este caso es conocida también como la Cedente, porque es la compañía que cede riesgo.

Las condiciones del negocio entre estas compañías quedan establecidas y reglamentadas en dichos contratos que a su vez están sujetos a la jurisdicción de la CNSF. Las cláusulas que los componen se enfocan en tres diferentes aspectos:

- actuariales o técnicos, que rigen la parte económica de la relación;
- administrativos, que especifican tiempos y formas de los flujos y documentos; y
- jurídicos, que aclaran procedimientos, vigencia, legalidad y sustento del mismo.

Dentro de las cláusulas jurídicas destaca la cláusula de *errores u omisiones*, la cual tiene su fundamento en el concepto de la *buena fe* y establece en general que cualquier error u omisión cometido por alguna de las partes no necesariamente implica una rescisión del contrato, siempre y cuando éste sea corregido una vez que se hubiese detectado.

Es la existencia de esta cláusula uno de los fundamentos que permitieron y originaron el proyecto al que hago mención en este documento y por el cual se motiva su mención.

### 1.3 Objetivos y limitaciones del reaseguro

Cada empresa puede tener particularidades sobre lo que busca obtener a través del reaseguro, tanto por las diferentes operaciones de los seguros<sup>1</sup> como por sus productos o mercados específicos; sin embargo, en términos generales se puede decir que el reaseguro:

- Minimiza las fluctuaciones en el pago de reclamaciones.
- Aumenta la capacidad para emitir montos superiores de suma asegurada.
- Coopera en los márgenes de solvencia financiera para una compañía.
- Forma parte de la estrategia para diversificar el riesgo asumido.
- Proporciona experiencia (la de la reaseguradora) en la selección de riesgos<sup>2</sup>.
- Proporciona información de productos, administración, reclamaciones o canales de distribución (ventas) de los mercados internacionales a la Cedente<sup>2</sup>.

Técnicamente existen varias formas de reaseguro, cada una de ellas cumple con alguno o algunos de los objetivos arriba mencionados, por lo que una compañía aseguradora puede conseguir varios de estos objetivos al mismo tiempo por medio de los diferentes tipos de reaseguro disponibles.

Sin embargo, este enfoque de compartir o ceder el riesgo presenta ciertas limitaciones:

- ◆ No puede transformar un riesgo no asegurable en uno asegurable.
- ◆ No puede convertir un riesgo mal investigado o mal suscrito<sup>3</sup> en uno bien suscrito.
- ◆ Ayuda a limitar la gravedad de las pérdidas, mas no acota la frecuencia de las mismas.
- ◆ No transfiere o elimina la responsabilidad legal para con los asegurados.
- ◆ Utilizado en exceso puede provocar un daño financiero a la Cedente.

---

<sup>1</sup> Existen tres agrupaciones diferentes para las operaciones de los seguros, que son: vida, daños y accidentes y enfermedades.

<sup>2</sup> Una reaseguradora se caracteriza por ser una empresa internacional, dado que necesita diversificar los riesgos asumidos.

<sup>3</sup> La suscripción es un proceso de clasificación del riesgo real de los asegurados previo a la emisión de sus pólizas.

## 1.4 Cúmulos de riesgo

Para una compañía de seguros es necesario contar con una buena medición de los riesgos asumidos. Las pólizas emitidas representan las obligaciones legales adquiridas; sin embargo, por diversas razones, en algunos de los ramos de operación de los seguros<sup>4</sup> pueden existir diferentes pólizas que protegen un mismo riesgo y es en ese tipo de situaciones que se dan los llamados cúmulos de riesgo.

Al menos en el ramo de vida individual<sup>4</sup>, una persona puede tener acceso a los diferentes productos que cubren necesidades de protección, educación o ahorro y por lo tanto puede comprar varias pólizas que finalmente implican para la compañía una acumulación de los riesgos amparados. Esto es claro al pensar que en caso de siniestro, la indemnización a pagar a los beneficiarios es la suma de las responsabilidades de cada una de las pólizas del asegurado.

Desde el punto de vista de la medición de los riesgos reales, en este ramo, cuando un mismo asegurado contrata más de una póliza, las sumas aseguradas de cada una de ellas deben sumarse para obtener el riesgo total que el asegurado ha transferido a la compañía.

De esta manera, las cesiones a los reaseguradores deben basarse en los cúmulos de riesgo, pues estos representan el riesgo total real aceptado y es el monto que está sujeto a la supervisión y normatividad de las entidades regulatorias.

El presente trabajo está enfocado al ramo de vida individual<sup>5</sup>, por lo que hablaremos de cúmulos como si se tratara del conjunto de pólizas de un mismo asegurado; por facilidad en el lenguaje, a menudo se utiliza la palabra cúmulo como sinónimo de la palabra asegurado.

## 1.5 Problemas en la identificación de cúmulos

Por varios motivos, en la cartera de vida de una compañía de seguros pueden existir problemas en la identificación de los cúmulos de riesgo, que al final son prácticamente problemas por no poder identificar de manera única a cada cliente.

El negocio de seguros existe desde tiempos remotos y aunque presenta una formalización ya en los inicios del siglo XX, es hasta fechas muy cercanas al año 2000 cuando la industria (y en general todo el mundo) comienza a ocuparse en desarrollar metodologías, procesos y sistemas de identificación única para sus clientes.

Como vida individual es el único ramo que desarrolla sus productos para darles seguimiento y atención a largo plazo (más de un año), hereda registros diversos de sus clientes desde varias décadas atrás, sin la garantía de que estén identificados plenamente. Es decir, si algún asegurado con pólizas antiguas obtiene alguna nueva, éste puede no ser identificado como cliente preexistente en la compañía. Además, los cambios en las tecnologías de información empleadas, donde se almacenan los datos de los asegurados, no necesariamente son compatibles entre sí.

---

<sup>4</sup> Las operaciones de los seguros se dividen a su vez en ramos; para la operación de vida existen los ramos individual, grupo y colectivo y pensiones.

<sup>5</sup> Las actividades profesionales del autor en la compañía donde labora se ocupan únicamente del ramo de vida individual.



Por otro lado, modificaciones en políticas internas y hasta la presencia del error humano en la captura de datos, son factores influyentes para que no se puedan identificar adecuadamente los cúmulos de vida.

Las implicaciones más importantes en las que puede caer una compañía de seguros por no realizar de manera correcta los cúmulos para reaseguro son los siguientes:

- Sobrepasar la capacidad máxima de retención de un riesgo autorizada por la CNSF.
- Desviaciones en los resultados esperados de la operación de reaseguro por:
  - Una incorrecta cesión del riesgo y el pago de una menor e injusta prima de reaseguro en los contratos por excedentes.
  - Insuficiente recuperación de montos en caso de siniestros.
  - Duplicidad en las cesiones de riesgos en contratos tipo cuota parte.
- Por los tres puntos anteriores se puede caer en alguna violación a dichos contratos.
- Por lo tanto, se pueden endurecer las negociaciones con los reaseguradores.

Como se puede apreciar, el tema de los cúmulos de riesgo tiene su relevancia especial en las compañías de seguros, más aún cuando estas situaciones no se deben sólo a errores humanos propios de la captura de los datos en la emisión de las pólizas, sino a un conjunto de circunstancias ajenas a la culpa de alguna persona.

## Reseña

Este reporte de trabajo pretende describir y señalar toda la labor realizada en el proyecto de la *estandarización de nombres* de la cartera de vida individual de la institución de seguros a la que pertenece el autor. El proyecto se realizó desde febrero de 2005 hasta mediados del 2006, no continuamente, dado que la carga de trabajo requirió la dedicación a otras actividades ajenas al proyecto, pero que en cada una de las etapas del mismo el autor estuvo presente y aportó sus conocimientos y capacidad de análisis para la realización del mismo, participando desde la planeación del proceso a seguir en el proyecto, incluyendo el desarrollo de un algoritmo probabilístico para la identificación de palabras coincidentes, hasta las conclusiones finales que apoyaron la toma de decisiones de la empresa.

## 2. Descubrimiento de los errores en los cúmulos de vida individual

### 2.1 El origen de la controversia

Durante el ejercicio de 2003, se suscitaron varios casos de siniestro que, por su monto en suma asegurada, rebasaron el límite de retención que la empresa de seguros mantenía en los contratos automáticos por excedentes que manejaba en aquel momento<sup>6</sup>.

El personal encargado de la administración de reclamos dio aviso de los casos al área de la administración del reaseguro para que se efectuara la recuperación procedente; sin embargo, éstos últimos se percataron de que los asegurados no estaban registrados como personas que cedieran riesgos al contrato automático de reaseguro.

Lo anterior ocurría porque los asegurados involucrados poseían varias pólizas, donde su nombre no había sido capturado en el sistema de la misma manera, por lo que el sistema no los había identificado para acumular sus sumas aseguradas y ceder en reaseguro el riesgo excedente al límite de retención.

Por medio de la cláusula de errores y omisiones del contrato se procedió en los primeros casos al intercambio de prima adeudada por la cedente en los años no cedidos del riesgo y la recuperación de los siniestros, cuyo resultado es a favor de la cedente en todos los casos.

La frecuencia de estos siniestros acaecidos en el mismo año obligó a la reaseguradora a solicitar audiencia con el área técnica para dialogar sobre este problema, pues al parecer habría muchos casos como éstos y por lo mismo, la compañía debería ceder más riesgos al reasegurador, abonando todas las primas adeudadas, además de atender lo estipulado en la cláusula de errores y omisiones donde se exige reparar el error una vez que se ha detectado.

### 2.2 Extensión del tema a toda la cartera

Presentado el problema de los cúmulos no identificados, dimos inicio, como área técnica, a la investigación e implicaciones del tema, en sus causas, orígenes y consecuencias.

El área de sistemas informó que el espacio reservado para el nombre del asegurado en la base de datos era un vector de 35 caracteres, sin restricciones ni validaciones. El sistema contaba además con un identificador de cliente (un número secuencial asignado) y un campo para el RFC de los asegurados, pero que éste había sido creado apenas 10 años atrás y aunque se validaba el formato del mismo, no se cotejaba con el nombre capturado.

El "id" del cliente en cualquier caso hacía referencia al nombre del asegurado o era capturado en caso de que el cliente lo supiera, por lo que no garantizaba la unicidad de los asegurados.

Además, los avances tecnológicos no se actualizaron a la par, particularmente en el uso de caracteres propios del español. Es decir, mientras el sistema de emisión permitía la letra "ñ" y los acentos, el sistema administrador de pólizas no, interpretando los caracteres raros como pudiera (no precisamos cómo se hubiese dado esto a niveles de lenguajes de computación).

---

<sup>6</sup> Los contratos automáticos obligan al reasegurador a aceptar todo riesgo dentro de los parámetros económicos y de suscripción definidos en el mismo.

Por otro lado, la recién creada área de normatividad y servicio no tenía fundamentos de políticas específicas para la captura de nombres, aunque el personal operativo contaba con manuales de prácticas que trataban de seguir, sin que esto significara una garantía.

Los operadores y capturistas comentaron que los formatos y las solicitudes de pólizas han ido cambiando e incluso han estado vigentes varias al mismo tiempo<sup>7</sup>. Esto les afecta porque en unas reciben nombre y apellido por separado y en otras en un solo renglón. Además de que se han enfrentado a confusiones por letra poco legible, abreviaturas, títulos, etc.

Otro aspecto es que las políticas o prácticas les han sido comunicadas paulatinamente, por lo que las pólizas han sido capturadas bajo lineamientos diferentes a través del tiempo. Por ejemplo, actualmente se les exige anotar los nombres con un formato específico y rechazar solicitudes ilegibles, confusas o con datos faltantes, mientras que antes no había indicaciones al respecto.

Por otra parte, el área técnica y taller de productos dictaminaron lineamientos particulares para ciertos planes: el seguro de educación, donde se capturan datos tanto de los padres como del hijo y los seguros de vidas conjuntas, donde dos personas están protegidas.

Cada caso tiene sus particularidades: por ejemplo, el seguro educacional pide anotar el nombre del hijo en el campo del nombre del asegurado y el del padre en el del contratante<sup>8</sup>, incluso este producto diseñó su propia solicitud para la toma de datos de ambas personas.

Para vidas conjuntas se consideraron tablas de edad equivalentes para efectos de la prima a cobrar, desligando del sistema el cálculo de edad desde la fecha de nacimiento. Al no dictaminar una política sobre la captura de este dato, pólizas de este tipo presentan cualquiera de las dos fechas de nacimiento: la que eligiera el capturista. Con respecto al nombre, se usó la práctica de que en alguna parte del campo se escribiera una "y" para vincular los dos nombres involucrados en la póliza, pero ésta podía estar en cualquier lugar; aunado esto al límite de caracteres, los operadores tenían que usar su creatividad para llenar el requisito, recurriendo al uso de ingeniosas abreviaturas o el dejar palabras truncas.

En la investigación también se dedujo que de manera natural algunos asegurados podrían llenar sus solicitudes con nombres diferentes, por actualización de su profesión o estado civil, sin que ello derive de algún error en sí. Tal situación fue confirmada por los agentes.

Por último, la presencia del error humano está presente en todo momento: los asegurados tal vez anotaron mal su fecha de nacimiento, con faltas de ortografía su nombre o letra ilegible; el agente u operador pudieron tener confusiones visuales en las pantallas de captura (como escribir el nombre en otro campo), cambiar nombres por apellidos o los típicos errores de tecleo; los programadores fallar en los procesos de mantenimiento ocasionales, haciendo que se sustituyera algún nombre con caracteres raros; etcétera.

Por todo lo anterior, se dimensionó que el dilema se extendía a toda la cartera, trascendiendo al reaseguro, tanto automático como facultativo<sup>9</sup>, e incluyendo a las pólizas terminadas que, en caso de rehabilitación, caerían eventualmente en el error de una acumulación incorrecta.

---

<sup>7</sup> Esto también ocurre por la diversidad de productos, que ha venido incrementándose y donde algunos requieren más información de los clientes.

<sup>8</sup> Todas las pólizas requieren de un contratante, que es la persona física o moral que se responsabiliza del pago de primas. Puede ser el mismo asegurado.

## 2.3 Causales de error en nombre

Después de analizar las bases de asegurados de diversas fuentes, las consultas con varias áreas de la compañía y cierto análisis deductivo, se clasificaron las causas de los errores en la comparación de nombres y se sugirieron algunas soluciones que apoyaron el diseño posterior de todo el proceso para identificar a los asegurados iguales con nombres diferentes.

La siguiente tabla preliminar muestra una clasificación con comentarios y estatus de cada causa particular identificada, en una etapa previa al proceso definitivo:

Causa de error	Ejemplo	Solución propuesta	Comentarios (previos)
1. Diferente cantidad o tipo de caracteres que separan las palabras	LUZ ANAYA LUZ ANAYA LUZ – ANAYA	Separar palabras eliminando los caracteres “separadores”.	Se programó una macro que separa las palabras correctamente.
2. Orden en que fueron capturadas las palabras	LUIS UGALDE UGALDE, LUIS	Hacer búsquedas por palabras separadas	Se generaron las tablas, falta programar las búsquedas
3. Caracteres extraños dentro de las palabras	NU%EZ RODR&GUEZ	Identificar casos y cambiarlos manualmente	Hay un total de 397 casos, hace falta cambiarlos
4. Truncamiento de palabras	GUILLER, MARGARI	Identificar casos y cambiarlos manualmente	Se pueden encontrar con una lista de todas las palabras separadas.
5. Abreviaturas	MA. FCO. GPE. GZL. GMO.	Eliminar abreviaturas Utilizar proyecto del área del CRM	Su proceso no está listo. Eliminar las que se vean en la lista de todas las palabras separadas.
6. Faltas de ortografía o errores de tecleo	GARSA, VRENDIA, FKAVIO, LIUS	Utilizar proyecto del área del CRM	Su proceso no está listo. Sustituir identificando casos con la lista de las palabras separadas.
7. Palabras juntas	DELEON, ALMAMARIA	POR DEFINIR	Difícilmente se puede hacer un algoritmo, existen apellidos así.
8. Mayúsculas, minúsculas y acentos	LUA, gerardo GONZALES, alma johana	Identificar casos y sustituir las letras que correspondan.	Casos 100% hallados: No hay acentos. Sólo hay 8 nombres con minúsculas.
9. Escritura de N en lugar de Ñ	ACUNA, YANEZ	Dejar en todos los casos la letra N	No existe la Ñ en el sistema

Después de arduas revisiones, ya que se tenía definido un proceso a seguir (que se describe en siguiente capítulo), se encontraron además los siguientes tipos de errores:

- ◆ **Direcciones como nombre:** A nuestro entender, las causas de estas inconsistencias se atribuyeron a imprecisiones en la captura o confusión del campo en el sistema. Otra posibilidad pudo ser algún mantenimiento del sistema. Algunos ejemplos:

- HIDALGO SUR 140 SALTILLO C                      EDUARDO MOLINA INTERIOR 1
- RICARDO FLORES MAGÓN 54                      GUERRERO RIVAPALACIO 13 MTY

<sup>9</sup> El reaseguro facultativo aplica para riesgos altos que salieron de los parámetros del automático y requieren negociación directa con los reaseguradores.

- ◆ Nombres de personas morales<sup>10</sup>: Error curioso y detectado en pocas pólizas. Se supuso causado por confusión en la captura del dato (que debió ser anotado en el campo del contratante) o por alguna indebida transacción del sistema al momento de la captura, pues en varios casos el contratante era un nombre de asegurado (persona física) pero en la mayoría se repetía la persona moral. Ejemplos:
  - CONSULTORES TOFRA, S.C. CHECKPOINT INTERNATIONAL S.A. DE CV
  - DISTRIBUIDORA TAMARIZ, S.A. FARMACIA HOMEOPATICA EL MANANTIAL
- ◆ Caracteres extraños: Este tipo de error se atribuyó directamente al sistema, pues sólo se detectaron 4 cadenas de caracteres extraños y en 3 de ellas se repetía esta en un bloque de varias pólizas:
  - 01 00000D10000100 N NN NNNNNNNN,NNNNN
  - ----- R0w3 &, 44444 /1
- ◆ Pólizas con el nombre nulo: Encontramos unos cuantos registros de pólizas con el nombre de asegurado en blanco, sin indagar sobre su posible causa.

Para los anteriores errores se buscó el nombre del asegurado en el campo del contratante y en los propios expedientes de las pólizas; las búsquedas en expedientes fueron un recurso utilizado como última instancia para autenticar la personalidad de los asegurados dudosos o averiguar la de los desconocidos, como en la mayoría de los casos anteriores.

Sin embargo, hubo otras situaciones de error que no requirieron de esta solución, pero que por las necesidades del algoritmo diseñado necesitaron un tratamiento especial. Este tipo de casos se listan enseguida:

- ◆ RFC después del nombre: Para estos casos se contaba con el RFC escrito también en su espacio, por lo que para efectos del proceso definitivo se solicitó eliminarlo del nombre. Los casos detectados fueron los siguientes:
  - DEVORA NUNEZ,ANGEL DENA520725-9AA
  - GERARDO RAMIREZ RUIZ RARG050579
- ◆ Consideraciones especiales para ciertos asegurados:
  - RODRIGUEZ DUE-AS,CONSTANTINO Cambiar a N el guión
  - ROMERO Y SALDA-A,ENRIQUE Cambiar a N el guión
  - GRA,EN MAYNEZ,MANUEL Cambiar a N la primer coma
  - BARONA RIVERA,BEATRIZ XAMAN-EE No separar el guión
  - THIERRY MTZ VULLIEZ-S,JEAN YVES No separar el guión

Considerar como casos especiales a estos asegurados derivó de la definición hecha para separar palabras en uno de los pasos del proceso definitivo, donde los caracteres “coma” y “guión” entre otros, se eliminaban por considerarlos caracteres que delimitan palabras.

A continuación se presenta el proceso definido para la identificación de errores en cúmulos.

<sup>10</sup> Es incorrecto que una empresa o razón social esté sujeta al riesgo de fallecimiento.

### 3. Proceso de identificación de cúmulos no detectados

#### 3.1 Objetivos y alcances

Básicamente lo que se busca con este proceso es identificar a los asegurados que tienen diversas pólizas, conociendo cuales son todas y cada una de ellas, incluyendo las pólizas terminadas<sup>11</sup>. Esto último para tener antecedente de todos los asegurados que ha tenido la empresa y para contemplar los casos de posibles rehabilitaciones<sup>12</sup>.

Sin embargo no estaba dentro de los alcances del proyecto la corrección de los nombres o la sustitución de los mismos, por diversas razones (fiscales, de contratación, del propio sistema, desconocimiento del nombre real y completo de los asegurados, etc.).

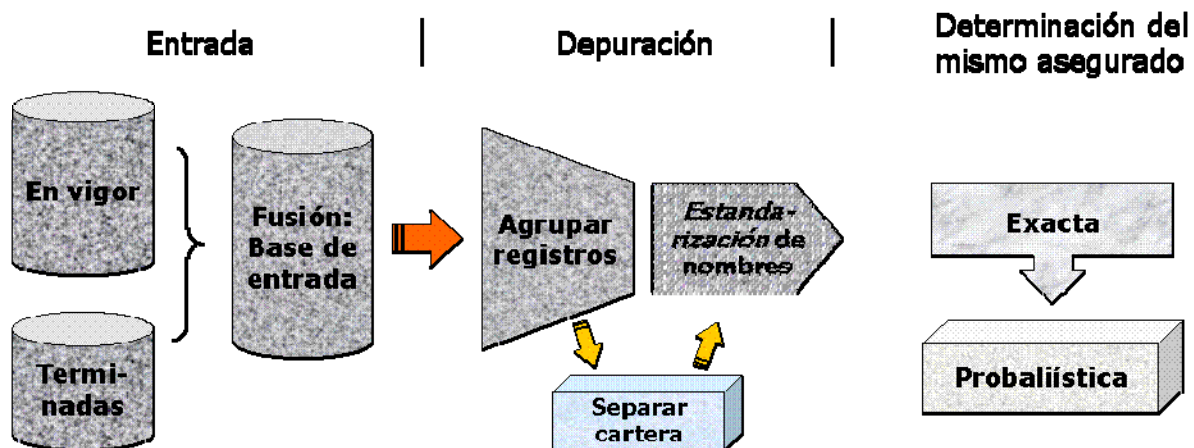
El enfoque de los resultados serviría para resolver la problemática del reaseguro en primera instancia, aunque el trabajo sentaría precedente para considerar cúmulos en nuevas emisiones para evitar futuros errores ("detener la bola de nieve"), donde el personal de emisión revisaría que cualquier nuevo asegurado efectivamente sea "nuevo". Además se respaldarían los resultados para su uso con fines estadísticos, como por ejemplo para saber cuantas pólizas promedio tiene un mismo asegurado.

Por último cabe mencionar que el cómo dejar los resultados, hablando en términos de formatos, archivos, bases de datos, responsables de mantenimiento y demás, también quedó fuera del alcance de este proyecto, esperando que se resolviera la situación del reaseguro.

#### 3.2 Diseño general

Como se presentó en el capítulo anterior, con la investigación e información recabada, se diagnosticaron varios problemas y causas del porqué la cartera presentaba asegurados cuyo riesgo total no podía identificarse en el sistema de reaseguro. Analizando las soluciones que habría que dar a cada parte del problema, se diseñó de manera completa el proceso mediante el cual la compañía podría identificar a los asegurados en esta situación.

El siguiente es el esquema general del proceso desarrollado:



<sup>11</sup> Razones para que una póliza esté terminada: fin de vigencia, expiración natural, siniestro, falta de pago y rescate.

<sup>12</sup> Los asegurados pueden reinstalar la vigencia de sus pólizas, pagando las primas adeudadas cuando ésta se cancela por falta de pago.

El proceso fue dividido en tres fases: entrada, depuración y determinación del mismo asegurado. La fase de entrada debe generar una base de datos con la información principal útil de las pólizas, como lo son el número de póliza, el nombre del asegurado, la fecha de nacimiento, la edad de emisión y la edad real. Esta fase tiene sus detallas, que más adelante se explican, debido a las características y circunstancias de las fuentes de información.

En la fase de depuración se reducen los registros de la base de datos de entrada por medio de una agrupación por llave y/o separación de cartera<sup>13</sup> y se trabaja con el nombre del asegurado de cada póliza para estandarizarlo o normalizarlo, generando un *nuevo nombre* con el que se puedan hacer comparaciones con el respectivo *nuevo nombre* de las demás pólizas en la siguiente fase, y se logre determinar así si asegurados de pólizas diferentes son el mismo o no.

### 3.3 El problema con las fuentes de información

El sistema administrador almacena de manera separada varios grupos de información. Existen archivos exclusivos que guardan los datos de los asegurados, otros para los agentes, productos, pólizas, fondos financieros, factores de reservas, de valor garantizado, etc.

Aunque nuestro interés se centra en algunos pocos datos, requerimos acceder tanto a información de pólizas como de asegurados; dentro de los procesos de la compañía no existen como tal bases de datos que combinen la información requerida de esta manera.

Existe por ejemplo, el proceso de valuación de reservas, que genera mensualmente archivos confiables con respecto al vigor de la cartera, pero sin datos de los asegurados. También contamos cada mes con el archivo general de reaseguro, donde se refleja casi todo el vigor, salvo algunas excepciones, y que relaciona los números de póliza con los nombres de los asegurados, pero no contiene algunos otros datos de las pólizas (como la edad de emisión).

Además, ningún proceso genera extracciones con datos de las pólizas terminadas; para esto habría que generar una extracción particular en nuestro proyecto.

Las fuentes de información de pólizas, entonces, son dos: un archivo maestro de las pólizas en vigor y otro donde quedan registros de las pólizas terminadas. En ambos casos, los datos utilizados para este proceso de identificación de cúmulos son: el número de póliza, suma asegurada, edad de emisión, edad real, fecha de emisión, clave de plan y número de cliente.

Los datos del asegurado requeridos son: nombre, fecha de nacimiento, dirección (que está almacenada en 4 campos; se utilizan todos), contratante, RFC y, claro, el número de cliente.

Para probar los resultados de este proyecto se aprovechó la existencia de un par de bases de pólizas (una de vigor y otra de terminadas) que se habían generado recientemente para otro propósito de reaseguro y que contenían la información necesaria, aunque no toda la cartera, quedando pendiente el diseño de un subprograma que extrajera la información de manera óptima y generara las bases de entrada ya con toda la cartera.

---

<sup>13</sup> Omitiendo un grupo de pólizas que comparten ciertas características en común, como el pertenecer a un mismo plan o a un mismo año de emisión.

### 3.4 Fase 1. Base de datos de entrada

Para lograr los objetivos mencionados, el proceso se realizaría dos veces, la primera vez sólo con la base de vigor y la segunda corrida con la fusión de las bases de vigor y terminadas.

Los resultados de la primera corrida servirían para solucionar las situaciones de las pólizas reaseguradas al momento del corte empleado, y los de la segunda nos darían luz para tomar acciones preventivas en casos de rehabilitaciones y dejar una base de consulta de nombres estandarizados histórica, misma que podría ocupar el área de suscripción para evitar futuros cúmulos desconocidos<sup>14</sup>.

Para el caso de la segunda corrida en esta fase, un programa simple amplía la base de pólizas en vigor añadiendo los registros de las pólizas terminadas, verificando previamente que no se repitieran registros al hacerlo y agregando un identificador de estatus a cada registro para mantener un buen control de las pólizas.

Cabe mencionar que en esta fase quedó programado un procedimiento que, para cuando el proceso se ejecutara con otras bases de pólizas (por ejemplo para toda la cartera), devolviera un nuevo archivo, donde, además de los campos con los datos de las pólizas que se tuvieran en dichas bases, se incluyeran los campos:

- ◆ Número de cliente
- ◆ Nombre (texto de 35 caracteres)
- ◆ Fecha de nacimiento
- ◆ Dirección (4 campos en total)
- ◆ RFC válido (no nulo y con el formato oficial)
- ◆ Contratante
- ◆ Edad de emisión (si no estuviera en la base)
- ◆ Edad real (si no estuviera en la base)

El archivo resultante se procesa en las siguientes fases del procedimiento.

### 3.5 Fase 2. Depuración de la información

Esta fase consta de una primera subrutina de agrupación para mejorar la eficiencia del proceso, otra opcional para separar bloques de pólizas que, en principio, tiene la misma intención de eficiencia, y el procedimiento principal que permite a la siguiente fase comparar a los asegurados, bautizado como el proceso de estandarización de nombres.

#### 3.5.1 Agrupación:

Mediante la concatenación del número de póliza con el nombre del asegurado y la fecha de nacimiento, se forma la llave con la que se agrupan los registros de la base, logrando reducir la cantidad de datos a analizar.

---

<sup>14</sup> El personal de suscripción podría buscar los nombres de asegurados nuevos considerando normas de estandarización (sólo para efectos de la búsqueda en la base de nombres histórica que generaríamos) a fin de encontrar pólizas anteriores y suscribir los riesgos de manera más adecuada.



Lo anterior se debe a que la información de las pólizas no tiene por qué llegar a la base de entrada como registros únicos, debido a varias características de los planes y de las pólizas como el que una misma póliza maneje más de una cobertura, más de un asegurado o que por sistema se administre en dos o tres fases.

La programación empleada simplemente elimina los registros que por llave resultan iguales.

### 3.5.2 Separación de cartera:

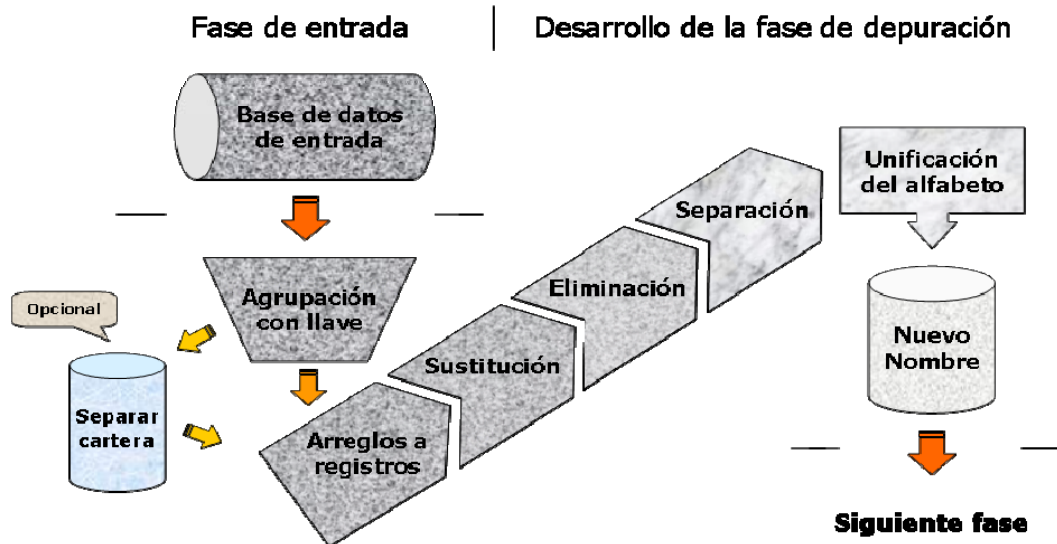
En los inicios del diseño se prefería omitir segmentos de pólizas con el fin de reducir los tiempos, como por ejemplo las pólizas del reaseguro facultativo, por tratarse de casos suscritos específicamente con cada reasegurador<sup>15</sup>, o las carteras de vidas conjuntas y del producto educacional, por las particularidades que ya se mencionaron en el capítulo anterior; sin embargo, como pueden participar en cúmulos de otras pólizas al final se decidió dejarlas.

La cartera que sí se separó de los análisis fue aquella que no cuenta con un nombre válido y que no se encontró alguno al revisar los expedientes de las pólizas, generalmente por falta del expediente mismo, aunque su exclusión en realidad no redujo el tiempo de los programas ya que estos casos representaron menos del 0.05% del total a analizar.

El motivo de su omisión sólo es lógico, pues no se puede trabajar con pólizas cuyo asegurado es desconocido y la forma de hacerlo fue a través de un listado de pólizas.

### 3.5.3 Flujo de información:

El siguiente esquema ilustra el paso de los datos a través de los módulos involucrados en esta fase:



El camino que se observa desde los "arreglos a registros" hasta la formación del "nuevo nombre" conforman las subrutinas del *proceso de estandarización de nombres*.

<sup>15</sup> Los asegurados con pólizas facultativas se suscribieron uno a uno con los distintos reaseguradores, no necesariamente participantes del contrato automático, donde se detectó el problema. Además estas pólizas se administran en otro servidor, por lo que se pensó sensato no involucrarlas de inicio.

### 3.5.4 Proceso de estandarización de nombres:

El primer módulo realiza algunos arreglos necesarios a ciertos registros. Por medio de claves de plan y criterios de fechas de emisión que se dieron al equipo de sistemas, a la cartera del seguro de educación se le sustituye el nombre del asegurado por el nombre del contratante, así como consideraciones específicas de la edad y fecha de nacimiento<sup>16</sup>. De esta manera, estas pólizas quedan alineadas al resto para poderse comparar.

Después, se genera una copia del campo “nombre del asegurado”, para no perder el nombre original del sistema, y en ella se corrigen los casos especiales detectados en las revisiones previas (que incluyeron a los contratantes del seguro educacional), mediante una tabla en la que se describen las acciones a realizar para cada póliza, como en los ejemplos del capítulo anterior al final del tema “causales de error en nombre” y los puntos 7 y 8 de la tabla ahí mostrada.

Concluido el módulo de arreglos, se sustituyen los caracteres extraños por la letra N; única situación detectada para la presencia de estos caracteres, después de los casos especiales. Se proporcionó la siguiente tabla con los caracteres a sustituir y su código ASCII:

código	35	38	47	96	166	188	196	207	208	239
caracter	#	&	/	`		¼	Ä	Ï	Ð	Ï

Los cambios se hacen sobre la copia del nombre original. Los caracteres fueron encontrados en los análisis previos ya descritos en el capítulo anterior.

A continuación, el proceso elimina las cadenas de caracteres que se enlistan en la tabla de eliminación de palabras construida para tal fin. La intención es desechar los títulos y cargos que se hayan capturado, ya que estas palabras no son parte del nombre de una persona.

En los albores del proyecto sólo se pensaba eliminar los cargos, títulos y sus abreviaturas comunes, como licenciado (lic.), arquitecto (arq.), general (gral.), etc. Sin embargo, al examinar los casos se encontraron abreviaturas que complicaban el flujo del proceso, como la del contador público o las referentes a maestrías académicas, por hallarse escritas con espacios entre sus iniciales<sup>17</sup>. Además, se detectaron posibles problemas en la ejecución, ya que algunas de las abreviaturas podían aparecer dentro de palabras completas o, en algunas ocasiones, sus terminaciones coincidían con el inicio de un nombre, y en consecuencia, su eliminación borraría la primera letra del nombre.

Por lo tanto, la construcción de la tabla y las programaciones propuestas sufrieron varias modificaciones, hasta que se decidió incluir en la tabla definitiva cadenas de caracteres y no palabras en sí, con una columna especificando el tamaño de la cadena a eliminar, cuidando e incluyendo las situaciones donde el espacio en blanco delimitaba el inicio o fin de la palabra o abreviatura que nos interesara.

De esta manera, sólo procede la eliminación si la coincidencia es exacta y se aplica sobre la copia del nombre del asegurado que ya ha sido modificado por los anteriores módulos.

<sup>16</sup> Se verificó previamente que esto fuera posible: el campo del contratante y del nombre del asegurado tienen el mismo tamaño y los mismos atributos.

<sup>17</sup> Situación no necesariamente catalogada como error, pero sí derivada de la falta de normas en la captura de los nombres de los asegurados.

A continuación se muestra un extracto de la tabla de eliminación de “palabras”:

Cadena	Longitud	Cadena	Longitud	Cadena	Longitud	Cadena	Longitud
ACTUARIA	8	LICENCIADA	10	, C P	6	C.P.A.	7
ACTUARIO	8	LICENCIADO	10	C P.	5	C.P.	5
AGRONOMO	8	LIC	5	,C P	5	,C.P.A.	7
AGRO	6	LIC.	5	,C P.	5	,C.P.	5
,AGRO	6	,LIC	5	C P A	7	PFRA	6
AGRO.	6	,LIC.	5	C P A.	7	PFRA.	6
,AGRO	6	GENERAL	7	,C P A	7	PROF	6
ING. IND.	10	GRAL	6	,C P A.	7	PROF.	6
,ING. IND.	10	,GRAL	6	L C P	7	PROFESOR	8
ING.IND	9	GRAL.	6	L C P.	7	PROFESORA	9
ING.IND.	9	,GRAL.	6	,L C P	7	DON	5
,ING IND	9	PADRE	7	,L C P.	7	,DON	5

La tabla completa contiene 1,220 cadenas de caracteres distintas. Cabe mencionar que se dejó la instrucción de que el espacio “ ” inicial o final de las cadenas no aplique cuando se compara con el inicio o término del campo donde está el nombre. Esto evita tener que generar una tabla mucho más grande.

El siguiente paso es el de la separación de palabras, donde el nombre de los asegurados se divide en las palabras que lo conforman, en campos independientes, con la intención de hacer búsquedas de coincidencias de palabras (en la siguiente fase) y mantener un estándar de palabras entre sí.

Por las investigaciones previas se supo que los caracteres espacio “ ”, coma “,” y guión “-” son los que delimitan las palabras. Esa es la razón para ejecutar la eliminación de cadenas de caracteres antes de la separación de los nombres en palabras.

Para cada registro, el nombre modificado hasta ahora es separado palabra por palabra y cada una genera su propio campo. Las palabras divididas no deben contener caracteres definidos como separadores de palabras, salvo los dos casos donde se solicitó no separar el guión, citados en el capítulo anterior.

Con las consideraciones anteriores, las palabras resultantes permiten únicamente caracteres de la A a la Z, el par de guiones mencionado y el punto “.”, pero aún quedan algunas abreviaturas de nombres que no se eliminaron, aunque serán sustituidas más adelante por la palabra completa.

Hasta el momento ya no tenemos espacios, comas, números, letras minúsculas ni caracteres extraños. El tema de los acentos no representó problema alguno, puesto que los sistemas aparentemente nunca los permitieron o siempre lograron sustituirlos por las letras respectivas sin acento.

En la antesala de la generación del nuevo nombre viene el módulo de sustitución de palabras y abreviaturas restantes, identificado en el diagrama como “unificación del alfabeto”, ya que en ese sentido el resultado es que se reduce y uniforma el universo de palabras a comparar en la siguiente fase.

Para lo anterior, se creó una tabla de 230 palabras o, nuevamente, cadenas de caracteres, que se sustituyen por la palabra que la misma lista define. Esta tabla es menos extensa que la de eliminación en parte porque las búsquedas no se hacen sobre un campo completo, donde la cadena puede encontrarse en cualquier parte, sino que es una búsqueda palabra por palabra en cada uno de los campos que se crearon, por lo que no se necesita abordar tantos casos; es decir, en la tabla de eliminación se ocupan varias cadenas de caracteres distintas para encontrar una misma palabra y aquí esto no es necesario.

A continuación se muestra un extracto de la tabla de sustitución de “palabras”:

Palabra	Sustitución	Situación	Palabra	Sustitución	Situación
BARRRERA	BARRERA	Error de tecleo	ENRIQUES	ENRIQUEZ	Unificar alfabeto
CABALLE	CABALLERO	Palabra trunca	ENRRIQUEZ	ENRIQUEZ	Error de tecleo
CABALLER	CABALLERO	Palabra trunca	ESCAREQO	ESCARENO	Reemplazo de Ñ
CAMAREN	CAMARENA	Palabra trunca	FDA	FERNANDA	Abreviatura
CASILLAS	CASILLAS	Error de tecleo	FNDA	FERNANDA	Abreviatura
CASTAQEDA	CASTANEDA	Reemplazo de Ñ	FDEZ	FERNANDEZ	Abreviatura
CEBALL	CEBALLOS	Palabra trunca	FDZ	FERNANDEZ	Abreviatura
CEBALLO	CEBALLOS	Palabra trunca	FERNADEZ	FERNANDEZ	Error de tecleo
CECEQA	CECENA	Reemplazo de Ñ	FERNANDE	FERNANDEZ	Palabra trunca
CONSEPCION	CONCEPCION	Unificar alfabeto	FNDEZ	FERNANDEZ	Abreviatura
CONTRERAZ	CONTRERAS	Unificar alfabeto	FDO	FERNANDO	Abreviatura
DUEQAS	DUENAS	Reemplazo de Ñ	FERNADO	FERNANDO	Error de tecleo

En la columna “situación” se plasman los motivos que dieron origen a la sustitución de las palabras, según nuestras deducciones, con lo que fue posible indicar especificaciones en la programación para algunas, que en total son cinco:

- Abreviatura: Referente a los casos donde intencionalmente una palabra se acortó, posiblemente por el problema del espacio. Para esta categoría, en la programación de las búsquedas quedó establecido que se buscaran adicionalmente estas palabras con un caracter punto “.” al final de la misma, con lo que se disminuyeron los registros totales de la tabla.
- Error de tecleo: Clasificación de las palabras en las que se nota alguna letra de más, faltante o cambiada y algunos errores ortográficos (los muy evidentes).
- Palabra trunca: Palabras que evidentemente fueron recortadas. Situación atribuida a la falta de espacio en la captura de los nombres.
- Reemplazo de Ñ: En estos casos se dedujo que el carácter Ñ fue cambiado por otra letra, tal vez de manera consciente o interpretado por algún sistema. Casi todos los casos se encontraron con la “Q” pero hubo también palabras con “P” y con “O” como MUPOZ y NUOEZ. La deducción se hizo plausible porque en estos casos de “P” u “O” también hubo al menos otra con “Q”.
- Unificación del alfabeto: Podemos decir que los errores ortográficos quedan aquí englobados, aunque no necesariamente algún cambio se deba a ortografía, pues sabemos que en cuestión de nombres, cada quien lo escribe como está en su acta.

La construcción de la tabla vino de la observación de una especie de histograma de palabras con sus frecuencias, en las revisiones previas, obedeciendo al sentido común y la sensatez del autor (con la venia de los jefes) tanto al definir las palabras a corregir como al establecer las que quedarán en sustitución.

Finalmente, el último módulo del proceso reescribe el nombre del asegurado en el campo que dio origen a la separación de palabras, una por una, separadas por un solo espacio y escritas conforme al resultado del paso anterior.

Así pues, lo que se tiene al final de esta fase es un archivo con los campos:

- ◆ Número de cliente
- ◆ Nombre (texto de 35 caracteres)
- ◆ **Nuevo nombre**
- ◆ Fecha de nacimiento
- ◆ Dirección (4 campos en total)
- ◆ RFC válido (no nulo y con el formato oficial)
- ◆ Contratante
- ◆ Edad de emisión
- ◆ Edad real
- ◆ **Grupo de campos de palabras** (con las palabras del nuevo nombre)

### 3.6 Fase 3. Determinación del mismo asegurado

Esta fase se forma de 2 algoritmos principales, como se aprecia en el diseño general. El primero de ellos, bastante sencillo, es el que determina de manera exacta al mismo asegurado, y el segundo, más complejo, lo hace de forma probable. Ambos se basan en la comparación de ciertos datos de cada registro contra la información análoga del resto de ellos, identificando los que cumplan con los criterios de validación definidos en cada uno.

#### 3.6.1 Algoritmo de manera exacta:

De inicio, el algoritmo genera una columna numérica adicional que denomina *nuevo número de cliente*. Luego, localiza las pólizas donde se tiene la seguridad de que sus asegurados son la misma persona y asigna a los registros involucrados un mismo número que resulta ser consecutivo en cuanto al conteo de casos que cumplen con alguno de los siguientes criterios:

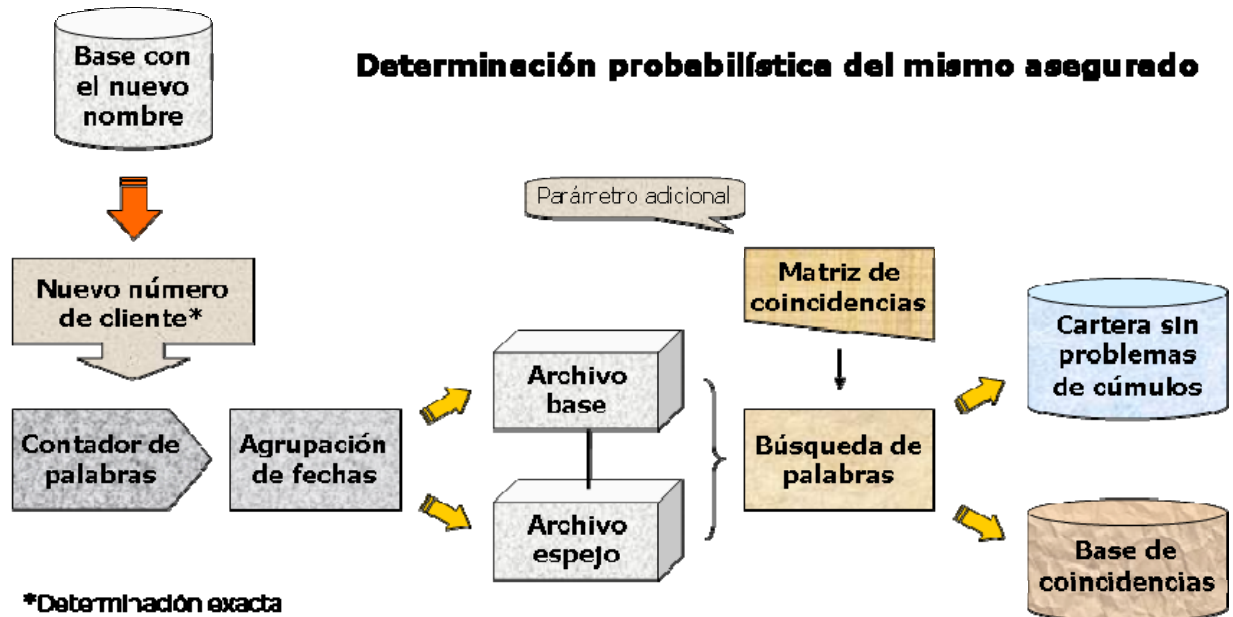
- ◆ Nuevo nombre y fecha de nacimiento iguales
- ◆ Fecha de nacimiento y RFC (válido) iguales
- ◆ Fecha de nacimiento y dirección iguales (los 4 campos)
- ◆ Mismo número de cliente original

Los resultados guardan consistencia al mantener a los asegurados que con anterioridad se sabe son el mismo con un número igual entre ellos.

### 3.6.2 Algoritmo probabilístico:

El algoritmo final realiza lo que fue la idea inicial para encontrar la solución a este problema enorme: la búsqueda cruzada de palabras. La idea surgió del conocimiento de pólizas cuyos nombres estaban anotados en orden diferente; uno de los siniestros más severos que dieron origen al tema fue el de una persona con dos pólizas, donde en una de ellas el nombre se anotó como Carlos Roberto y en la otra Roberto Carlos.

La siguiente ilustración muestra el algoritmo gráficamente:



Primeramente es necesario preparar bien la información para robustecer el algoritmo. Como se observa, éste comienza con un sencillo módulo contador de palabras, que añade otro campo numérico a la base, en donde se anota para cada registro el número de palabras en que fue dividido el nuevo nombre. El saber cuántas palabras tiene cada registro es fundamental para establecer un criterio de discriminación entre los nombres que pueden llegar a ser de la misma persona, ya que en la búsqueda cruzada de palabras se contarán cuántas palabras coinciden entre los nombres.

La idea original contemplaba hacer las búsquedas cruzadas de palabras entre todos los registros de la base uno por uno, lo cual es descabellado desde el punto de vista de los recursos del sistema y tiempo de ejecución, por lo que se consideró lógico limitar los cruces mediante la fecha de nacimiento. Naturalmente, es ocioso comparar nombres de personas que tiene gran diferencia en edad.

Con apoyo de ciertas estadísticas de las fechas de nacimiento, se estimó que el manejo de grupos de evaluación determinados por estas fechas, evitaría las comparaciones sin sentido antes mencionadas con una confianza del 97.61%<sup>18</sup>. Se anexa nota del estudio realizado.

<sup>18</sup> En otras palabras, el error de clasificación de las pólizas que influyera en la detección de cúmulos incorrectos se estima en 2.39%.

Dada la dificultad para validar las fechas de nacimiento de los posibles errores y por la reducción (en escala factorial) del número de operaciones que a nivel sistema se realizarían, se consideró prudente continuar bajo el esquema propuesto. De este modo, se construyó el módulo de agrupación de fechas en el que virtualmente se divide la base de datos, agrupando los registros que coinciden según la fecha de nacimiento.

El siguiente módulo genera una copia de la base de datos denominándola "archivo espejo", quedando la original como "archivo base" y diferenciando los títulos de las columnas para distinguirlos en el programa que ejecuta las búsquedas de palabras. Los campos del archivo base inician con "A." y los del espejo con "B."

Ambos archivos, junto con la matriz de coincidencias, alimentan al programa de la búsqueda cruzada de palabras, donde cada nombre del archivo base será evaluado contra todos los registros de su mismo grupo en el archivo espejo (a excepción de él mismo) de la siguiente manera:

- Se comparan todas las palabras de ambos nombres (una por una) contando de forma acumulada las que sean exactamente iguales entre sí, con los candados suficientes para no contar coincidencias dobles.
- La pareja de nombres es clasificada como probable si el número de palabras coincidentes es mayor o igual al determinado en la matriz de coincidencias.
- Hay que guardar en un archivo nuevo los registros completos (con todos los datos) de las parejas clasificadas como probables. El archivo debe tener su propio contador de renglones y una columna con la cantidad de palabras coincidentes bien identificada.

Al final de las comparaciones, las pólizas que no quedaron en ninguna de las dos columnas del archivo nuevo implícitamente forman la cartera sin problemas de cúmulos, y el archivo final del algoritmo es al que se le nombró base de coincidencias.

Dicha base fue sujeta a revisiones manuales para comprobar la eficacia del proceso.

### ***3.6.3 Matriz de coincidencias***

Es el clasificador de las comparaciones. Se trata de una matriz triangular de tamaño  $n$ , donde  $n$  es el máximo número de palabras encontradas en algún nombre. La primera columna indica, en orden ascendente, la cantidad de palabras del primer nombre, o nombre "A" y el primer renglón el número de palabras del segundo nombre (o nombre "B"); en el resto de sus componentes, la matriz indica el mínimo de palabras en que la pareja de nombres deben coincidir para que probablemente sean el mismo asegurado.

La matriz original se concibió por intuición, pensando aún en los cruces completos entre todos los registros, por lo que se buscó ser estrictos en la clasificación; sin embargo, al definir los grupos de evaluación anteriores, se relajó este criterio para probar, entre otros aspectos, los resultados del proceso (a través de revisiones manuales, como se ha citado) y encontrar para futuras corridas una matriz mucho más precisa.

A continuación se muestran las matrices descritas:

INICIAL	Nombre B									Asegurados
Nombre A	2	3	4	5	6	7	8	9	10	
2	2	2	2	2	2	2	2	2	2	176,246
3		2	2	3	3	3	3	3	3	583
4			3	3	4	4	4	4	4	77,201
5				3	4	4	5	5	5	74,104
6					4	5	5	6	6	13,442
7						4	5	6	6	8,378
8							5	6	6	1,902
9								5	6	492
10									6	135
										9

FINAL	Nombre B							
Nombre A	2	3	4	5	6	7	8	
2	2	2	2	2	2	2	2	2
3		2	2	3	3	3	3	3
4			3	3	3	3	3	3
5				3	3	3	3	3
6					3	3	3	3
7						3	3	3
8								3

La disminución de registros en la matriz final derivó de las depuraciones en la estandarización de nombres de la fase previa.

Esta fue la parte final del proceso computarizado; el archivo elaborado, como se mencionó con anterioridad (la base de coincidencias), fue revisado manualmente para verificar la eficacia del proceso completo. Las conclusiones, sin embargo, tuvieron que ver mucho más con decisiones estratégicas para la compañía que con la implantación definitiva del proceso como era la idea original, debido a que desde la primera ejecución se obtuvo información tan valiosa que modificó el rumbo de la solución al problema del que surgió este proyecto.



## 4. Resultados del proceso y conclusiones

### 4.1 Revisión de la base de coincidencias

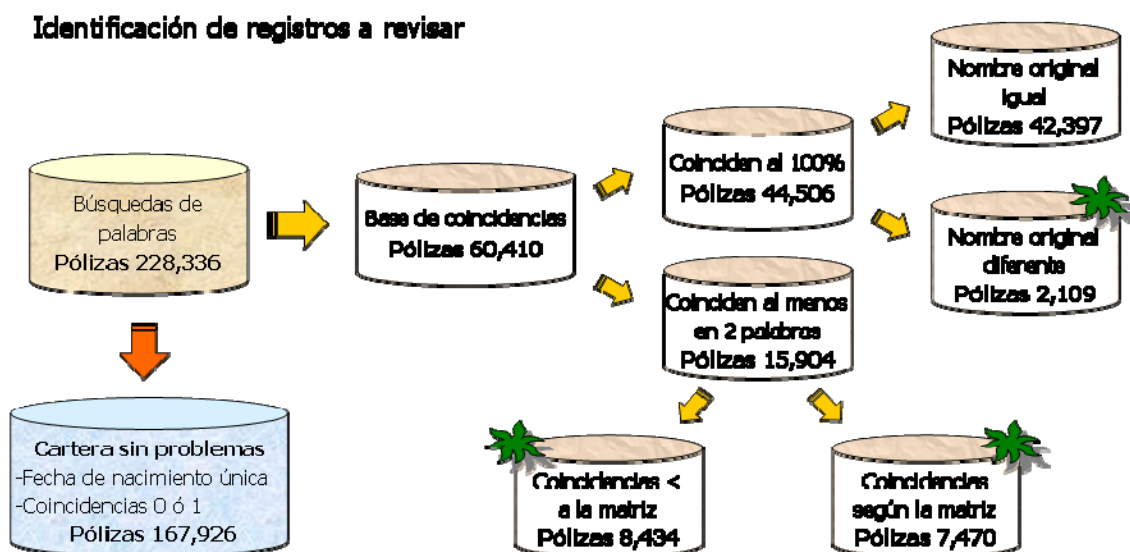
Las bases con las que se realizaron las búsquedas de palabras en la primera corrida del proceso contaron con 228,336 pólizas, ubicadas en 231,790 registros y conteniendo 199,957 nombres de asegurados a comparar. La base de coincidencias resultó quedarse con un total de 51,976 pólizas, mientras que las restantes 176,360 no tendrían problemas de cúmulos.

Para efectos de la revisión se clasificaron las pólizas restantes en tres tipos:

- Fecha de nacimiento única (pólizas con frecuencia de fecha de nacimiento igual a 1).
- Coincidencia de menos de 2 palabras (pólizas donde la búsqueda de palabras encontró intersección sólo en una o en cero palabras).
- Coincidencia de al menos 2 palabras (parejas de pólizas con 2 o más palabras iguales).

Esto fue con el fin de afinar bien los criterios de la matriz de coincidencias, pues en la tercera categoría podrían existir parejas de pólizas que por alguna razón (que también trataríamos de averiguar) fueran del mismo asegurado, a pesar de que teniendo más de 3 palabras en sus nombres estandarizados (según la matriz de coincidencias definitiva) coincidan sólo en dos.

De esta manera se agregaron los registros con dicha clasificación a la base de coincidencias, dividiéndola después en 4 partes, de acuerdo a la siguiente ilustración:



Se revisaron las bases marcadas, considerando las parejas de pólizas de cada una, sumando un total de 14,432 registros que contienen las 18,013 pólizas involucradas.

Con apoyo de personal operativo de la empresa se compararon uno por uno los registros, observando los datos disponibles de cada póliza, como los nombres (original y nuevo), RFC, dirección (4 campos), etcétera; distinguiendo el resultado de la comparación en colores tipo semáforo, donde el verde significa que la pareja de pólizas son del mismo asegurado, rojo que no es así, amarillo que se tiene duda y blanco si alguna póliza involucrada perteneciera a la cartera de vidas conjuntas, que como ya se ha mencionado, tiene importantes diferencias.

Los resultados de las revisiones a nivel registros fueron los siguientes<sup>19</sup>:

Coincidencia	Verde	Amarillo	Rojo	Blanco	Subtot
Al 100%	1,459	132	1	65	1,657
>= Matriz	2,772	238	1,286	1,499	5,795
< Matriz	120	20	6,402	438	6,980
Subtotales	4,351	390	7,689	2,002	14,432

Posteriormente se buscaron los expedientes de las pólizas dudosas para mejorar clasificación, reduciendo los casos amarillos, aunque no se pudieron eliminar todos por falta de algunos expedientes o información inexacta dentro del mismo:

Coincidencia	Verde	Amarillo	Rojo	Blanco	Subtot
Al 100%	1,588	2	2	65	1,657
>= Matriz	2,955	10	1,331	1,499	5,795
< Matriz	140	-	6,402	438	6,980
Subtotales	4,683	12	7,735	2,002	14,432

Para concluir de manera directa sobre los resultados de clasificación del proceso, se elaboró la siguiente estadística, omitiendo por supuesto, la cartera de vidas conjuntas:

Coincidencia	Al 100%	>= Matriz	< Matriz
Verde	99.7%	68.8%	2.1%
Amarillo	0.1%	0.2%	0.0%
Rojo	0.1%	31.0%	97.9%

Como se observa, la credibilidad de las bases de coincidencias al 100% y menor a la matriz (en esta ejecución: parejas de pólizas de tan sólo 2 palabras iguales) son muy aceptables.

Como se mencionó con anterioridad, en esta corrida se relajó el criterio de discriminación de la matriz de coincidencias. Hacer una estadística de las combinaciones de números de palabras por nombre en los resultados de la base de coincidencias mayores o iguales a los de la matriz, permitiría robustecerla, buscando alcanzar niveles de confiabilidad óptimos entre las secciones de *menor a* y *al menos* la matriz (al mover la matriz de coincidencias se puede perder efectividad en la base de coincidencias menores a la matriz).

Por otro lado, dentro de cada grupo encontramos además los registros que compartían ya un mismo "nuevo número de cliente", siendo consistentes en todos los casos con lo obtenido en el subproceso de determinación exacta del mismo asegurado, comprobando su eficacia.

## 4.2 Errores detectados en la fecha de nacimiento

Explorando las pólizas de los registros verdes, se detectó que algunos nombres cruzaron con la cartera que había sido clasificada como "sin problemas de cúmulos", principalmente con los de fecha de nacimiento única. Se decidió profundizar al respecto encontrando lo siguiente:

<sup>19</sup> Para medir la eficiencia del proceso interesa saber la proporción de parejas de pólizas que se identificaron bien, lo cual implica contar registros.

Haciendo comparaciones de los relativamente pocos registros encontrados de esta manera, como se hizo en las revisiones visuales antes mencionadas, encontramos 181 pólizas más involucradas en cúmulos pero cuyo problema se encontraba en un error en la captura de la fecha de nacimiento.

Los casos fueron en su gran mayoría evidentes, por ejemplo el que estuvieran intercambiados los meses con los días o que hubiera un número consecutivo de mes o día entre las fechas de los asegurados que “curiosamente” se llamaban igual y vivían en el mismo domicilio.

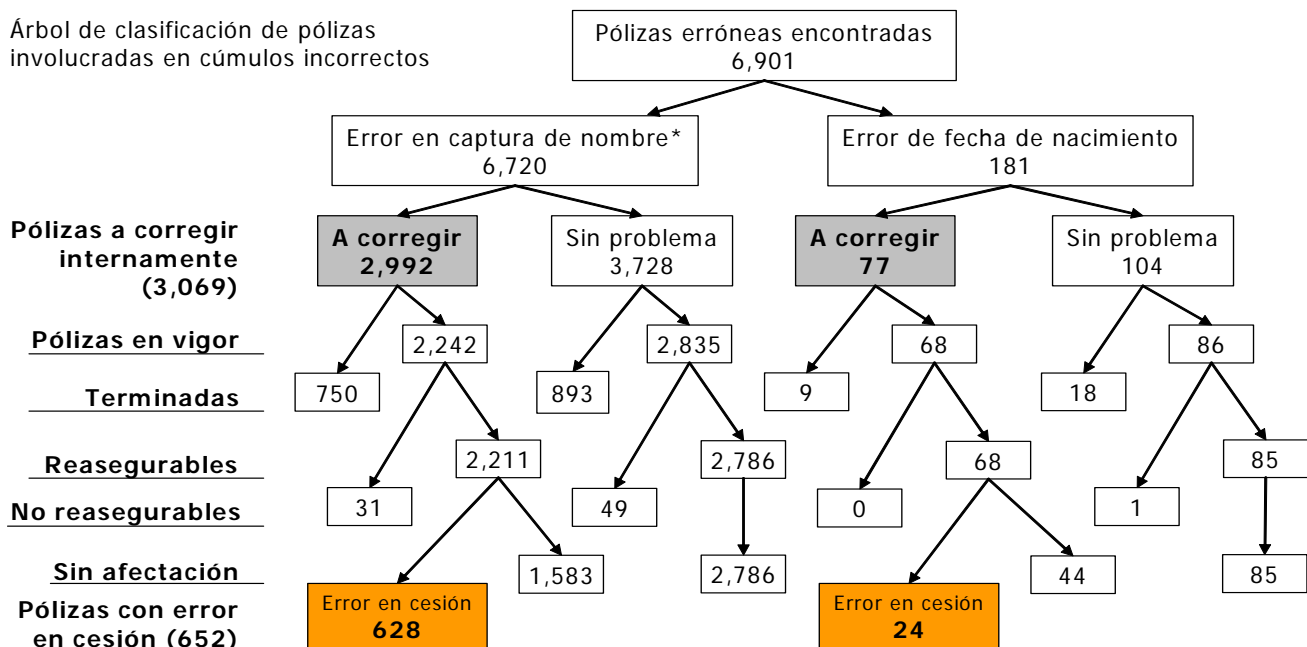
Juntando estas 181 pólizas con las 6,720 halladas en los 4,683 registros verdes sumamos un total de 6,901 pólizas involucradas en los errores de cúmulos.

### 4.3 Efectos de la corrección en cúmulos

Primero se verificó la situación actual en reaseguro de las pólizas coincidentes con nombre original igual, pues sus cúmulos debieran estar bien calculados. Quitando las intersecciones con las 6,901 pólizas anteriores, el estatus era correcto: sin errores en la cesión.

Ahora bien: la procedencia, estatus, tipo de producto y circunstancia específica de cada póliza tienen afectaciones diversas para que, a pesar de encontrarse en un cúmulo no realizado, éste no tenga afectación en el monto de cesión a reaseguro<sup>20</sup>. El análisis profundo de cada caso nos llevó a clasificar a las pólizas de acuerdo con el siguiente árbol:

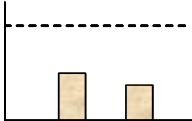
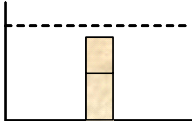
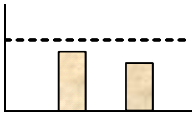
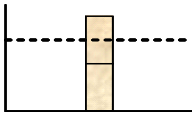
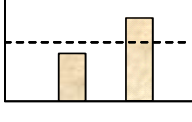
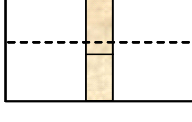
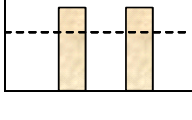
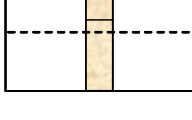
Árbol de clasificación de pólizas involucradas en cúmulos incorrectos



Como se observa, se identificaron 652 pólizas que, al incluirlas en los cúmulos que debieran pertenecer, afectan las cesiones de reaseguro.

<sup>20</sup> Por ejemplo, las pólizas cuyo nombre original era igual al estandarizado, o aquellas donde su monto en riesgo inicia un cúmulo pero éste no rebasa el límite de retención o ya se cedía por completo (al aumentar el cúmulo con una nueva póliza se continúa cediendo el mismo monto, sin cambios).

Para comprender mejor la forma en que cada caso afecta la cesión de reaseguro se diseñó el siguiente esquema explicativo, donde se aprecia cómo se modifican las cesiones de reaseguro con respecto al límite de retención al corregir los cúmulos. En él se cuenta de nuevo el total de pólizas involucradas y que se afectan por esto (sólo en vigor y reasegurables):

ANTES	DESPUÉS	Nombre del caso	Pólizas involucradas	Pólizas a corregir
		Sin implicaciones	4,108	0
		Nueva cesión	432	216
		Aumento de cesión	486	325
		Aumento de cesión y "baja" de asegurado	124	111
TOTALES:			5,150	652

Para tener un estimado previo sobre cuánto monto en riesgo y qué volumen de prima cedida representarían estas pólizas, se estudiaron varios escenarios de límite de retención, pues nuestros contratos permiten aumentarlo y nos encontrábamos en una etapa de transición<sup>21</sup>.

Al apreciar una volatilidad inesperada en los resultados, se profundizó en el funcionamiento y efecto del límite de retención para cada caso. Tomemos como ejemplo un cúmulo de tan sólo dos pólizas, donde la primera tiene un riesgo neto de cuatro millones de pesos y la segunda siete. Comenzando con un límite de retención de tres millones, la empresa tiene antes de corregir el cúmulo una retención total de seis millones para esta persona y una cesión de cinco. Al corregirlo, estas cantidades cambian a tres y ocho millones respectivamente, observando un aumento total de la cesión en los tres millones que se estaban reteniendo incorrectamente.

Ahora bien, si el límite de retención se encontrara en cuatro millones de pesos, la retención incorrecta sería de cuatro en lugar de tres. Ese monto va directo a cesión al momento de rectificar el cúmulo. La situación para un límite de cinco millones es parecida a la anterior, pues la compañía retiene erróneamente los cuatro millones de la primera póliza y por tanto la modificación en cesión vuelve a ser de cuatro millones de pesos.

<sup>21</sup> Además, un cambio de retención aplica sobre las pólizas en diferentes tiempos, de acuerdo a ciertas características especificadas en los contratos.

Por otro lado, si la retención máxima fuera de ocho millones, tendríamos que ninguna de las pólizas cedería riesgo a reaseguro por separado, pero al actualizar el cúmulo, éste sumaría once millones, de los cuales ocho se retendrían y el monto en error sería tan sólo de tres.

En resumen, en este ejemplo observamos cómo el monto retenido incorrectamente (igual al monto no cedido) y por ende, la prima adeudada (no pagada aún), aumentan hasta cierto nivel o tope, para disminuir a cero después, según el límite de retención que se aplique:

Límite de retención	3,000,000	4,000,000	6,000,000	8,000,000	10,000,000	11,000,000	12,000,000	15,000,000
Retención incorrecta	3,000,000	4,000,000	4,000,000	3,000,000	1,000,000	0	0	0
Efecto en la prima	aumenta	aumenta	constante	disminuye	disminuye	nula	nula	nula

El efecto en general para toda la cartera reasegurada es mucho más complejo. Aun así, después de una ardua investigación se logró percibir que al incrementar el límite de retención al doble, poco más del 95% de la cartera incorrecta caería en casos sin implicaciones para reaseguro, y aunque la aseguradora corriera un mayor riesgo por no ceder correctamente los pocos cúmulos remanentes, éstos cabían dentro de la retención autorizada, que es mucho mayor, lo cual no contraviene las disposiciones legales; además, la dispersión de los riesgos de toda la cartera permitían asumir dicho riesgo sin problemas.

Por otro lado, de diversas fuentes logramos constatar que el beneficio de tal operación superaba los costos asociados a la reparación de todos esos cúmulos (a nivel interno y externo, con los reaseguradores).

#### 4.4 Estrategia de reaseguro resultante

Las investigaciones y trabajos realizados gracias a la información obtenida de este proceso permitieron llegar a un acuerdo mucho más sencillo con los reaseguradores, evitar tediosas conciliaciones de cuentas y orientar la estrategia de reaseguro de la compañía hacia la obtención de un bajo costo neto de reaseguro<sup>22</sup>.

#### 4.5 Comentarios finales

El proyecto culminó de manera satisfactoria para las necesidades de reaseguro de la empresa de seguros; sin embargo, por el cambio de rumbo que siguió la estrategia y solución al embrollo desatado, quedaron al aire algunos temas u objetivos planteados o que emergieron durante la evolución de los análisis para el desarrollo del proceso.

Hasta el momento, no se ha ejecutado el proceso para toda la cartera, quedando pendiente la generación de una base de nombres general que el área de emisión y suscripción pudiera emplear en el reconocimiento de posibles nuevos cúmulos. Pretendiendo realizar esta acción, habría que actualizar o robustecer la matriz de coincidencias, tarea postergada por el camino alterno que siguió el desenlace del conflicto en reaseguro. Para tal efecto, se pueden retomar los resultados de la sección de cartera con coincidencias mayores o iguales a las de la matriz.

<sup>22</sup> Recordar que el reaseguro es un servicio de protección por el cual se paga un precio, siendo conceptualmente un costo para las compañías aseguradoras.

Aunado a lo anterior, a continuación presento las siguientes consideraciones que pudieran mejorar el desempeño del proceso:

- No sustituir las abreviaturas de nombres comunes como María (MA) si se encuentran como última palabra y sin punto, pues es más probable que se trate de un truncamiento.
- En el seguro de educación, el proceso pierde el nombre del niño. Una alternativa en este caso sería el intercambiar el nombre del asegurado con el del contratante, en lugar de que sólo se sustituya el del asegurado. Esto es válido a nivel sistema y no se alteran los nombres porque los tamaños de los campos son iguales.
- Una vez depurada la matriz de coincidencias, si se incluyen desde inicio del algoritmo probabilístico las fechas de nacimiento únicas, se pueden encontrar más posibles errores por esta causa.

#### **4.6 Experiencia personal**

Pues bien, el desarrollo de este proyecto en todas sus etapas dejó una vasta experiencia para el autor de este reporte, desde la investigación a detalle de aspectos tan sutiles en las necesidades y manejos que hay que dar a la información, pasando por las formas y fuentes de la misma, hasta la responsabilidad de llevar a cabo el proyecto, darle seguimiento, hacer que sucedan las cosas y obtener de él instrumentos y argumentos para generar estrategias encaminadas al provecho de la compañía.

Por otro lado, puedo identificar que, para la realización del proyecto, los conocimientos propios de la carrera fueron útiles en ciertos aspectos y necesarios en otros. La utilidad figura en el apoyo que me dieron para optimizar partes del proceso, como en los subprocesos para la determinación del mismo asegurado o en la creación de macros en Excel para realizar pruebas piloto que a su vez fueron guía de los algoritmos definidos al equipo de sistemas.

Otros apoyos que puedo considerar de mayor peso, fueron el análisis y la determinación de los valores para la matriz de coincidencias y el estudio de los resultados numéricos con los que se clasificaron las salidas de las ejecuciones del proceso, pudiendo obtener conclusiones sobre su efectividad. El posterior razonamiento de los efectos del límite de retención en los cúmulos de reaseguro, en cambio, exigió un dominio y conocimiento actuarial especializado.

En general, algunos momentos donde el conocimiento actuarial fue imprescindible, se dieron desde la comprensión de la problemática para diseño del proceso, la comunicación para su desarrollo y, como ya se adelantó un poco, para el análisis de resultados, cuya importancia radica en la aportación para la toma de decisiones.

Conocer los fundamentos del seguro, particularmente el de vida individual, su funcionamiento y comportamiento, se volvió elemental para explicar a otras áreas el cómo y porqué de varias dudas; el para qué se solicitaban ciertos campos en las extracciones; el orden en que se pensó el proceso y la importancia del proyecto por todo lo referente al reaseguro; las condiciones de los contratos; el aspecto regulatorio y del negocio de seguros de vida individual. De hecho, en el primer capítulo de este reporte se resumen todos estos conceptos.

El roce con otras áreas y el complemento de los conocimientos, sin embargo, son una de las lecciones más importantes que puede uno recolectar con estas experiencias de trabajo. Para muchos no es novedad el escuchar a otros referirse a la corriente educativa de continuidad en el aprendizaje, contrario a lo que uno puede llegar a pensar al terminar la carrera, o al menos, el imaginar lo que realmente significa, pues de acuerdo a nuestra experiencia en dicho momento, “no hay nivel de comparación” entre estudiar una carrera y el aprender otras cosas con el día a día. No obstante, complementar los conocimientos implica compartirlos y adaptarlos con más personas, actividad que no resulta tan sencilla.

El aprendizaje continuo toma relevancia cuando, en el desarrollo profesional, te percatas de que requieres también aprender a ser un líder, donde se involucra el desempeño de un equipo interdisciplinario de trabajo y por ende el de uno mismo. Este aprendizaje está también en un nivel de complejidad tan elevado, que pocos logran desarrollar.

Ampliar pues, los conocimientos en diversos aspectos del negocio del seguro, el contacto con las diferentes áreas de la compañía involucradas y el aprendizaje sobre trato y transformación de la información, son algunos de los legados, como ya lo he mencionado; sin embargo, uno de los más grandes logros que puedo citar por la realización de este trabajo, es sin duda el reconocimiento de los jefes, mismo que me ha llevado a ganar su confianza y fortalecer con ello mi crecimiento y desarrollo profesional, el cual ya ha rendido frutos.

Dentro de las más grandes satisfacciones que puede uno tener, está el de la trascendencia. El saber que dejas tu mejor esfuerzo y dedicación sobre un proyecto, que te interesa, te divierte y te enseña a la vez; que de él se construyen enormes y provechosos resultados, y del que se te reconoce, te hace sentir y constatar que vas con el rumbo correcto.

## Anexo único - Estadísticas de fechas de nacimiento

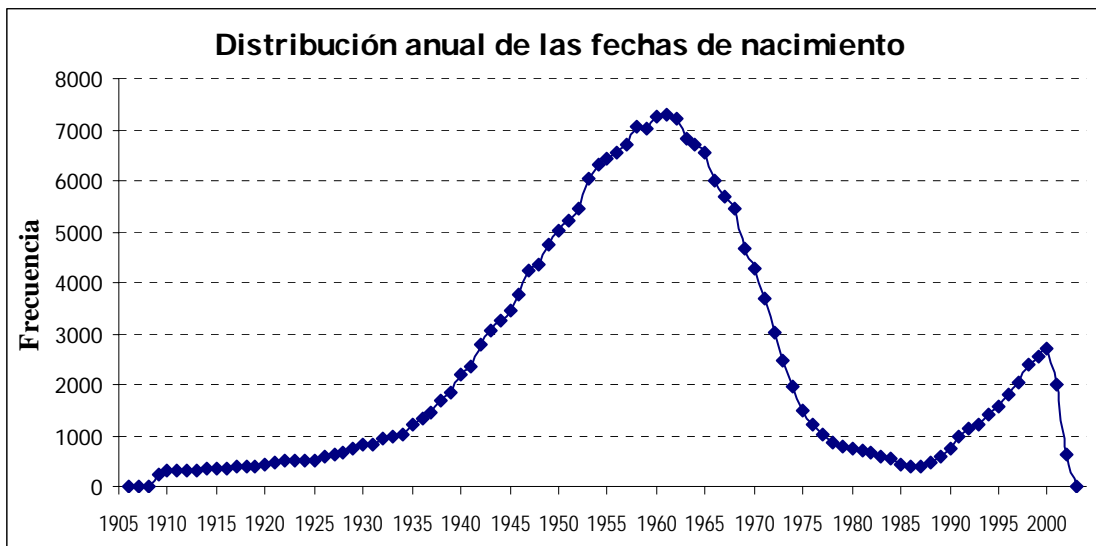
La base de pólizas en vigor utilizada para la primera ejecución del proceso de cúmulos no detectados tuvo su origen en una extracción que se solicitó anteriormente para otro proyecto de reaseguro. Comparando esta base contra las de valuación de reservas, que son la fuente más confiable del vigor con datos que este ejercicio requería para la misma fecha de corte, obtuvimos una cobertura de pólizas del 99.93%, de un volumen de casi 230,000, por lo que las conclusiones se consideraron válidas para toda la cartera.

En las bases de valuación podemos encontrar datos de la póliza como la fecha de emisión y los años de vigencia, así como las edades real y de emisión del asegurado\*.

Con los datos de la valuación calculamos la edad real actual de cada asegurado para compararla con la edad que resulta al considerar la fecha de nacimiento de la base original. Como lo muestra el cuadro de la derecha, la conclusión obtenida es que el 97.56% de las pólizas guardan consistencia con ambos datos.

Resta de edades	Frecuencia	Porcentaje
> -3	28	0.01%
-2	9	0.00%
-1	520	0.23%
<b>0</b>	<b>222,859</b>	<b>97.56%</b>
1	4,664	2.04%
2	59	0.03%
> 3	294	0.13%

Por otro lado, se observó el siguiente comportamiento:



El incremento de la cola derecha se debe al auge que tiene el seguro de educación desde los años 90's.

Asimismo, de la base original, se contaron poco más de 30 mil fechas de nacimiento distintas, de las cuales el 17% resultaron ser únicas, con lo que se pudo suponer que cerca de 5,100 pólizas tienen a un único asegurado. Estas pólizas, a su vez, conforman el 2.25% del total.

Pensando que estas pólizas de asegurados únicos se encuentran uniformemente distribuidas lo largo de la cartera, entonces el 2.25% de las pólizas que no coincidieron en la validación anterior tampoco representarían problemas de cúmulos, ampliando nuestro margen de confianza obtenido anteriormente de 97.56% a 97.61%.

\* La edad real es precisamente la que tiene el asegurado al momento de la emisión de la póliza, la de emisión es con la que se calculan sus tarifas.